For more than 20 years, a full dump of all dblp records in our own XML format has been available as open data for download and reuse. These dump files have always been in high demand over the years (with 500+ downloads in February 2022 alone) and are used as a research dataset in numerous publications.

For quite some time now, we have been asked to provide a **full RDF dump** as well. Snapshots of the dblp XML file have been converted to RDF before by members of the community, and there are still a number of those RDF files available on the Internet. However, the problem with these snapshots is that they are usually not updated once they are created. Given its continuous curation by the dblp team which makes dblp such a "living" dataset, external files will be left severely out of sync with the curated dblp data stock, in some cases up to several years.
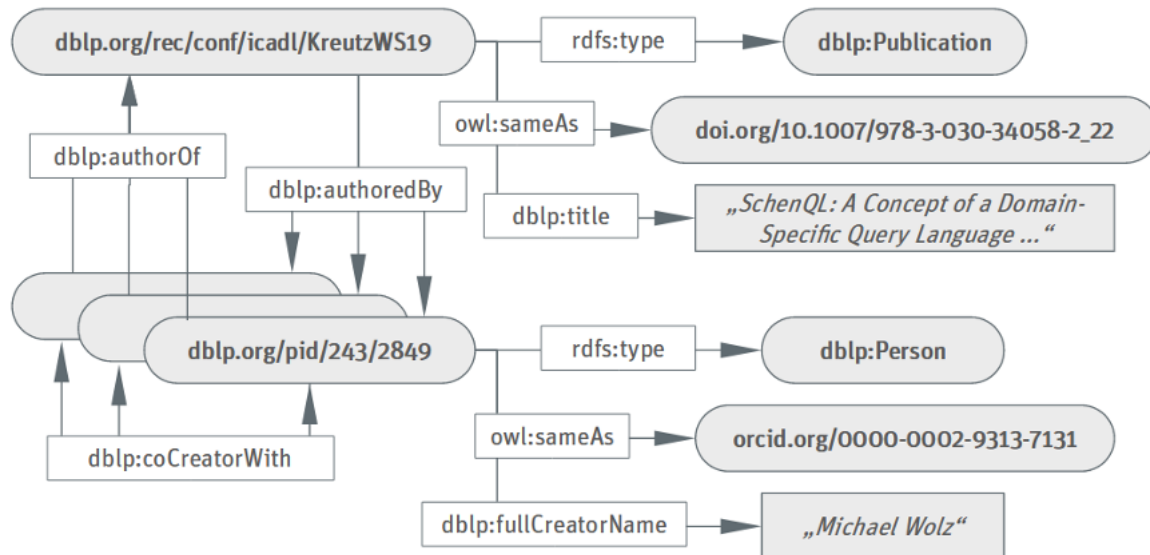
A full dblp RDF dump

To remedy this situation, we are happy to announce that for a few months now the entire dblp dataset has also been available as RDF data. These dump files will be updated daily and are guaranteed to always be in sync with the latest dblp XML release. Files are available as RDF/XML, N-Triples and Turtle at

- https://dblp.org/rdf/

However, if you plan to use our RDF dump in your experiments and you need a persistent snapshot in order to make your results reproducible, we strongly encourage you to use a **persistent monthly release** instead:

- https://dblp.org/rdf/release/

With its current schema, the dblp RDF data model forms a simple person-publication graph.

Example excerpt from the dblp RDF dump.

In total, the current release of March 2022 consists of 2,941,316 person entities, 6,010,605 publication entities, and 252,573,199 RDF triples. 12,157,035 external resource URIs are linked in the data set. Just as with any other data provided by dblp, the RDF dump is made available under CC0 1.0 Public Domain Dedication license.

RDF fragments via live API

Please be aware that besides the complete RDF dump, there also is the live API that serves RDF fragments in RDF/XML and N-Triples for individual entities. It can be used to query for persons or publications using their PID/key, e.g.:

- https://dblp.org/rec/conf/icadl/KreutzWS19.rdf
- https://dblp.org/rec/conf/icadl/KreutzWS19.nt

- https://dblp.org/pid/243/2849.rdf
- https://dblp.org/pid/243/2849.nt

Note that there is at least some rate limiting in place to protect the live API from

aggressive crawlers. That means, in case of a rather large number of queries, it is recommended to download the complete dump and do your queries locally.

Known limitations and future improvements

Numerous metadata aspects, like reference to the journal that an article is contained in or the affiliation of an author, are currently provided only as string literals. Future iterations of the schema will see these objects (i.e., publication venues and institutions/organizations) being added to the data model as true entities, together with their own metadata, persistent IDs and links to external resources. Hence, we don't see the dblp RDF schema as final, but rather as a first step in providing the semantics of the dblp dataset in a more structured way. We also aim to provide a proper SPARQL endpoint in the near future.

If you work with dblp RDF, we would of course be very grateful if you could share your thoughts, experiences, and criticisms with us. You can contact us as usual via email at dblp(at)dagstuhl.de, or send us your thoughts via Twitter @dblp_org. Your kind support of dblp is very much appreciated!

Acknowledgments

We'd like to thank the dblp Advisory Board as well as numerous individuals from the semantic web community for their ideas and comments regarding the beta version of the dblp RDF. Your feedback helped us to significantly improve our schema. We'd like to particularly thank Silvio Peroni, Hannah Bast, Ralf Schenkel and Tobias Zeimetz for their time for discussions, helpful criticisms, and concrete ideas that were crucial in developing the dblp RDF.